10.1 Confidence Intervals: Introduction TCC

Student Learning Outcomes

By the end of this chapter, the student should be able to:

- Calculate and interpret confidence intervals for one population mean and one population proportion.
- Interpret the student-t probability distribution as the sample size changes.
- Discriminate between problems applying the normal and the student-t distributions.

Introduction

Suppose you are trying to determine the mean rent of a two-bedroom apartment in your town. You might look in the classified section of the newspaper, write down several rents listed, and average them together. You would have obtained a point estimate of the true mean. If you are trying to determine the percent of times you make a basket when shooting a basketball, you might count the number of shots you make and divide that by the number of shots you attempted. In this case, you would have obtained a point estimate for the true proportion.

We use sample data to make generalizations about an unknown population. This part of statistics is called **inferential statistics**. **The sample data help us to make an estimate of a population parameter**. We realize that the point estimate is most likely not the exact value of the population parameter, but close to it. After calculating point estimates, we construct confidence intervals in which we believe the parameter lies.

In this chapter, you will learn to construct and interpret confidence intervals. You will also learn a new distribution, the Student's-t, and how it is used with these intervals. Throughout the chapter, it is important to keep in mind that the confidence interval is a random variable. It is the parameter that is fixed.

If you worked in the marketing department of an entertainment company, you might be interested in the mean number of digital songs a consumer streams per month. If so, you could conduct a survey and calculate the sample mean, x, and the sample standard deviation, s. You would use x to estimate the population mean and s to estimate the population standard deviation. The sample mean, x, is the **point estimate** for the population mean, μ. The sample standard deviation, s, is the point estimate for the population standard deviation, σ.

Each of x and s is also called a statistic.

A **confidence interval** is another type of estimate but, instead of being just one number, it is an interval of numbers. The interval of numbers is a range of values calculated from a given set of sample data. The confidence interval is likely to include an unknown population parameter.

Suppose for the song streaming example we do not know the population mean μ but we do know that the population standard deviation is $\sigma = 10$ and our sample size is 100. Then by the Central Limit Theorem, the standard deviation for the sample mean is

$\frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{100}} = 1.$

The **Empirical Rule**, which applies to bell-shaped distributions, says that in approximately 95% of the samples, the sample mean, x, will be within two standard deviations of the population mean μ. For our song streaming example, two standard deviations is $(2)(1) = 2$. The sample mean x is likely to be within 2 units of μ.

Because x is within 0.2 units of μ, which is unknown, then μ is likely to be within 0.2 units of x in 95% of the samples. The population mean μ is contained in an interval whose lower number is calculated by taking the sample mean and subtracting two standard deviations $((2)(0.1))$ and whose upper number is calculated by taking the sample mean and adding two standard deviations. In other words, μ is between $x - 0.2$ and $x + 0.2$ in 95% of all the samples.

For the song streaming example, suppose that a sample produced a sample mean $x = 20$. Then the unknown population mean μ is between

$x - 2 = 20 - 2 = 18$ and $x + 2 = 20 + 2 = 20$

We say that we are **95% confident** that the unknown population mean number of songs streamed per month is between 18 and 22. **The 95% confidence interval is (18, 22).**

The 95% confidence interval implies two possibilities. Either the interval (18, 22) contains the true mean μ or our sample produced an x that is not within 2 units of the true mean μ. The second possibility happens for only 5% of all the samples (100% - 95%).

Remember that a confidence interval is created for an unknown population parameter like the population mean, μ. Confidence intervals for some parameters have the form

(point estimate - margin of error, point estimate + margin of error)

The margin of error depends on the confidence level or percentage of confidence.

When you read newspapers and journals, some reports will use the phrase "margin of error." Other reports will not use that phrase, but include a confidence interval as the point estimate + or - the margin of error. These are two ways of expressing the same concept.

Note: Although the text only covers symmetric confidence intervals, there are non-symmetric confidence intervals (for example, a confidence interval for the standard deviation).

Optional Collaborative Classroom Activity

Have your instructor record the number of meals each student in your class eats out in a week. Assume that the standard deviation is known to be 3 meals. Construct an approximate 95% confidence interval for the true mean number of meals students eat out each week.

1. Calculate the sample mean.
2. $\sigma = 3$ and $n =$ the number of students surveyed.
3. Construct the interval $\left(x - 2 \cdot \frac{\sigma}{\sqrt{n}}, x + 2 \cdot \frac{\sigma}{\sqrt{n}}\right)$

We say we are approximately 95% confident that the true average number of meals that students eat out in a week is between __________ and ___________.

Glossary

Confidence Interval (CI)
: An interval estimate for an unknown population parameter. This depends on:

    - The desired confidence level.
    - Information that is known about the distribution (for example, known standard deviation).
    - The sample and its size.

Inferential Statistics
: Also called statistical inference or inductive statistics. This facet of statistics deals with estimating a population parameter based on a sample statistic. For example, if 4 out of the 100 calculators sampled are defective we might infer that 4 percent of the production is defective.

Parameter
: A numerical characteristic of the population.

Point Estimate
: A single number computed from a sample and used to estimate a population parameter.

10.2 Confidence Intervals: Confidence Interval, Single Population Mean, Population Standard Deviation Known, Normal TCC
Confidence Intervals: Confidence Interval, Single Population Mean, Population Standard Deviation Known, Normal is part of the collection col10555 written by Barbara Illowsky and Susan Dean with contributions from Roberta Bloom.

Calculating the Confidence Interval

To construct a confidence interval for a single unknown population mean μ **, where the population standard deviation is known,** we need x as an estimate for μ and we need the margin of error. Here, the margin of error is called the **error bound for a population mean** (abbreviated **EBM**). The sample mean x is the **point estimate** of the unknown population mean μ
The confidence interval estimate will have the form:

- (point estimate - error bound, point estimate + error bound) or, in symbols,$(x - \text{EBM}, x + \text{EBM})$

The margin of error depends on the **confidence level** (abbreviated **CL**). The confidence level is often considered the probability that the calculated confidence interval estimate will contain the true population parameter. However, it is more accurate to state that the confidence level is the percent of confidence intervals that contain the true population parameter when repeated samples are taken. Most often, it is the choice of the person constructing the confidence interval to choose a confidence level of 90% or higher because that person wants to be reasonably certain of his or her conclusions.

There is another probability called alpha (α). α is related to the confidence level CL. α is the probability that the interval does not contain the unknown population parameter.
Mathematically, α + CL = 1.

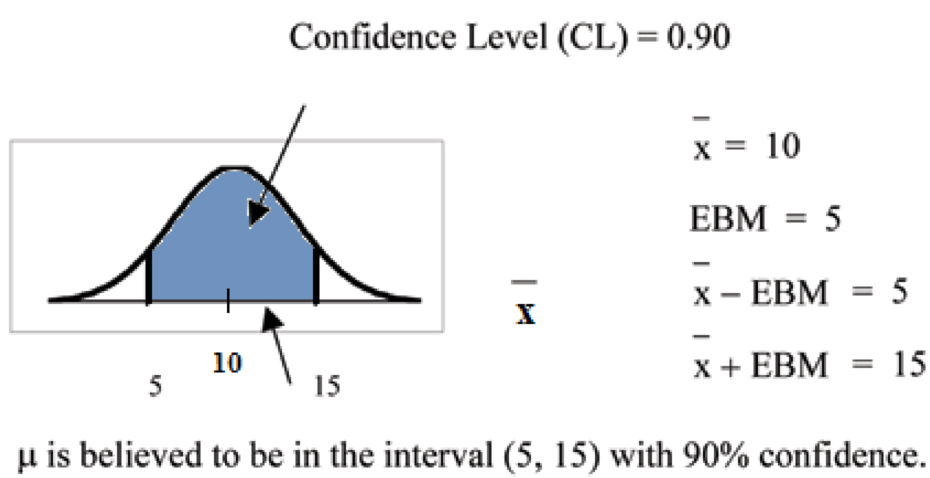
Example:

- Suppose we have collected data from a sample. We know the sample mean but we do not know the mean for the entire population.
- The sample mean is 7 and the error bound for the mean is 2.5.

$x = 7$ and $\text{EBM} = 2.5$.
The confidence interval is $(7 - 2.5, 7 + 2.5)$; calculating the values gives $(4.5, 9.5)$.
If the confidence level (CL) is 95%, then we say that "We estimate with 95% confidence that the true value of the population mean is between 4.5 and 9.5."

A confidence interval for a population mean with a known standard deviation is based on the fact that the sample means follow an approximately normal distribution. Suppose that our sample has a mean of $x = 10$ and we have constructed the 90% confidence interval (5, 15) where $\text{EBM} = 5$.

To get a 90% confidence interval, we must include the central 90% of the probability of the normal distribution. If we include the central 90%, we leave out a total of α = 10% in both tails, or 5% in each tail, of the normal distribution.



To capture the central 90%, we must go out 1.645 "standard deviations" on either side of the calculated sample mean. 1.645 is the z-score from a

Standard Normal probability distribution that puts an area of 0.90 in the center, an area of 0.05 in the far left tail, and an area of 0.05 in the far right tail.

It is important that the "standard deviation" used must be appropriate for the parameter we are estimating. So in this section, we need to use the standard deviation that applies to sample means, which is $\frac{\sigma}{\sqrt{n}}$. $\frac{\sigma}{\sqrt{n}}$ is commonly called the "standard error of the mean" in order to clearly distinguish the standard deviation for a mean from the population standard deviation σ.

In summary, as a result of the Central Limit Theorem:

- With a large sample size, X is normally distributed, that is, X ~ $\mathrm{N}\left(\mu_X, \frac{\sigma}{\sqrt{n}}\right)$.
- **When the population standard deviation σ is known, we use a Normal distribution to calculate the error bound.**

Calculating the Confidence Interval:

To construct a confidence interval estimate for an unknown population mean, we need data from a random sample. The steps to construct and interpret the confidence interval are:

- Calculate the sample mean x from the sample data. Remember, in this section, we already know the population standard deviation σ.
- Find the Z-score that corresponds to the confidence level.
- Calculate the error bound EBM
- Construct the confidence interval
- Write a sentence that interprets the estimate in the context of the situation in the problem. (Explain what the confidence interval means, in the words of the problem.)

We will first examine each step in more detail, and then illustrate the process with some examples.

Finding z for the stated Confidence Level

When we know the population standard deviation σ, we use a standard normal distribution to calculate the error bound EBM and construct the confidence interval. We need to find the value of z that puts an area equal to

the confidence level (in decimal form) in the middle of the standard normal distribution Z~N(0,1). You can use the following information for your corresponding z-score:

90% confidence level: z = 1.645

95% confidence level: z = 1.96

99% confidence level: z = 2.58

EBM: Error Bound
The error bound formula for an unknown population mean μ when the population standard deviation σ is known is

- $\mathrm{EBM} = z \cdot \frac{\sigma}{\sqrt{n}}$

Constructing the Confidence Interval

- The confidence interval estimate has the format $(x - \mathrm{EBM}, x + \mathrm{EBM})$.

Writing the Interpretation
The interpretation should clearly state the confidence level (CL), explain what population parameter is being estimated (here, a **population mean**), and should state the confidence interval (both endpoints). "We estimate with ___% confidence that the true population mean (include context of the problem) is between ___ and ___ (include appropriate units)."

Example:
Suppose scores on exams in statistics are normally distributed with an unknown population mean and a population standard deviation of 3 points. A random sample of 36 scores is taken and gives a sample mean (sample mean score) of 68. Find a confidence interval estimate for the population mean exam score (the mean score on all exams).
Exercise:

Problem:

Find a 90% confidence interval for the true (population) mean of statistics exam scores.

Solution:

- You can use technology to directly calculate the confidence interval
- The first solution is shown step-by-step (Solution A).
- The second solution uses the TI-83, 83+ and 84+ calculators (Solution B).

Solution A
To find the confidence interval, you need the sample mean, x, and the EBM.

- $x = 68$
- $\text{EBM} = z \cdot \left(\frac{\sigma}{\sqrt{n}}\right)$
- $\sigma = 3$; $n = 36$; The confidence level is 90% (CL=0.90)

$z = 1.645$

$$\text{EBM} = 1.645 \cdot \left(\frac{3}{\sqrt{36}}\right) = 0.8225$$

$$x - \text{EBM} = 68 - 0.8225 = 67.1775$$

$$x + \text{EBM} = 68 + 0.8225 = 68.8225$$

The 90% confidence interval is **(67.1775, 68.8225).**

Solution B
Using a function of the TI-83, TI-83+ or TI-84 calculators:

Press `STAT` and arrow over to `TESTS`.
Arrow down to `7:ZInterval`.

Press `ENTER`.
Arrow to `Stats` and press `ENTER`.
Arrow down and enter 3 for σ, 68 for x , 36 for n, and .90 for `C-level`.
Arrow down to `Calculate` and press `ENTER`.
The confidence interval is (to 3 decimal places) (67.178, 68.822).
Interpretation
We estimate with 90% confidence that the true population mean exam score for all statistics students is between 67.18 and 68.82.
Explanation of 90% Confidence Level
90% of all confidence intervals constructed in this way contain the true mean statistics exam score. For example, if we constructed 100 of these confidence intervals, we would expect 90 of them to contain the true population mean exam score.

Changing the Confidence Level or Sample Size

Example: Changing the Confidence Level
Exercise:

Problem:

Suppose we change the original problem by using a 95% confidence level. Find a 95% confidence interval for the true (population) mean statistics exam score.

Solution:

To find the confidence interval, you need the sample mean, x, and the EBM.

- $x = 68$
- $\text{EBM} = z \cdot \left(\frac{\sigma}{\sqrt{n}}\right)$
- $\sigma = 3$; $n = 36$; The confidence level is 95% (CL=0.95)

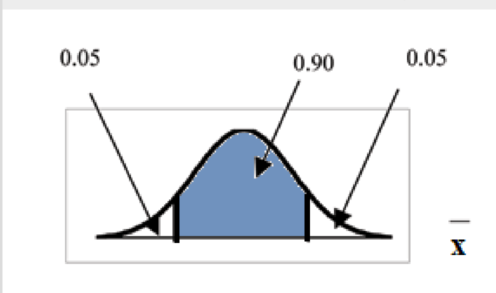
$$z = 1.96$$

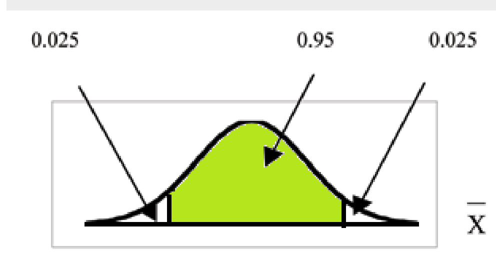
$$\text{EBM} = 1.96 \cdot \left(\frac{3}{\sqrt{36}}\right) = 0.98$$

$$x - \text{EBM} = 68 - 0.98 = 67.02$$

$$x + \text{EBM} = 68 + 0.98 = 68.98$$

Interpretation
We estimate with 95% confidence that the true population mean for all statistics exam scores is between 67.02 and 68.98.
Explanation of 95% Confidence Level
95% of all confidence intervals constructed in this way contain the true value of the population mean statistics exam score.
Comparing the results
The 90% confidence interval is (67.18, 68.82). The 95% confidence interval is (67.02, 68.98). The 95% confidence interval is wider. If you look at the graphs, because the area 0.95 is larger than the area 0.90, it makes sense that the 95% confidence interval is wider.





Summary: Effect of Changing the Confidence Level

- Increasing the confidence level increases the error bound, making the confidence interval wider.
- Decreasing the confidence level decreases the error bound, making the confidence interval narrower.

Example: Changing the Sample Size:

Suppose we change the original problem to see what happens to the error bound if the sample size is changed.

Exercise:

Problem:

Leave everything the same except the sample size. Use the original 90% confidence level. What happens to the error bound and the confidence interval if we increase the sample size and use n=100 instead of n=36? What happens if we decrease the sample size to n=25 instead of n=36?

- $x = 68$
- $\text{EBM} = z \cdot \left(\frac{\sigma}{\sqrt{n}}\right)$
- $\sigma = 3$; The confidence level is 90% (CL=0.90) ; $z = 1.645$

Solution:

If we **increase** the sample size n to 100, we **decrease** the error bound.

When $n = 100$: $\text{EBM} = z \cdot \left(\frac{\sigma}{\sqrt{n}}\right) = 1.645 \cdot \left(\frac{3}{\sqrt{100}}\right) = 0.4935$

Solution:

If we **decrease** the sample size n to 25, we **increase** the error bound.

When $n = 25$: $\text{EBM} = z \cdot \left(\frac{\sigma}{\sqrt{n}}\right) = 1.645 \cdot \left(\frac{3}{\sqrt{25}}\right) = 0.987$

Summary: Effect of Changing the Sample Size

- Increasing the sample size causes the error bound to decrease, making the confidence interval narrower.
- Decreasing the sample size causes the error bound to increase, making the confidence interval wider.

Working Backwards to Find the Error Bound or Sample Mean

Working Backwards to find the Error Bound or the Sample Mean
When we calculate a confidence interval, we find the sample mean and calculate the error bound and use them to calculate the confidence interval. But sometimes when we read statistical studies, the study may state the confidence interval only. If we know the confidence interval, we can work backwards to find both the error bound and the sample mean.

Finding the Error Bound

- From the upper value for the interval, subtract the sample mean
- OR, From the upper value for the interval, subtract the lower value. Then divide the difference by 2.

Finding the Sample Mean

- Subtract the error bound from the upper value of the confidence interval
- OR, Average the upper and lower endpoints of the confidence interval

Notice that there are two methods to perform each calculation. You can choose the method that is easier to use with the information you know.

Example:
Suppose we know that a confidence interval is **(67.18, 68.82)** and we want to find the error bound. We may know that the sample mean is 68. Or perhaps our source only gave the confidence interval and did not tell us the value of the the sample mean.
Calculate the Error Bound:

- If we know that the sample mean is 68: $\mathrm{EBM} = 68.82 - 68 = 0.82$
- If we don't know the sample mean: $\mathrm{EBM} = \frac{(68.82-67.18)}{2} = 0.82$

Calculate the Sample Mean:

- If we know the error bound: $x = 68.82 - 0.82 = 68$

- If we don't know the error bound: $x = \frac{(67.18+68.82)}{2} = 68$

Calculating the Sample Size n

If researchers desire a specific margin of error, then they can use the error bound formula to calculate the required sample size.

The error bound formula for a population mean when the population standard deviation is known is $\text{EBM} = z \cdot \left(\frac{\sigma}{\sqrt{n}}\right)$

The formula for sample size is $n = \frac{z^2\sigma^2}{\text{EBM}^2}$, found by solving the error bound formula for n

Example:
The population standard deviation for the age of Foothill College students is 15 years. If we want to be 95% confident that the sample mean age is within 2 years of the true population mean age of Foothill College students , how many randomly selected Foothill College students must be surveyed?

- From the problem, we know that $\sigma = 15$ and EBM=2
- $z = 1.96$, because the confidence level is 95%.

- $n = \frac{z^2\sigma^2}{\text{EBM}^2} = \frac{1.96^2 15^2}{2^2}$ =216.09 using the sample size equation.
- Use n = 217: Always round the answer UP to the next higher integer to ensure that the sample size is large enough.

Therefore, 217 Foothill College students should be surveyed in order to be 95% confident that we are within 2 years of the true population mean age of Foothill College students.

**With contributions from Roberta Bloom

Glossary

Confidence Interval (CI)

An interval estimate for an unknown population parameter. This depends on:

- The desired confidence level.
- Information that is known about the distribution (for example, known standard deviation).
- The sample and its size.

Confidence Level (CL)

The percent expression for the probability that the confidence interval contains the true population parameter. For example, if the $\mathrm{CL} = 90\%$, then in 90 out of 100 samples the interval estimate will enclose the true population parameter.

Error Bound for a Population Mean (EBM)

The margin of error. Depends on the confidence level, sample size, and known or estimated population standard deviation.

10.3 Confidence Intervals: Confidence Interval, Single Population Mean, Standard Deviation Unknown, Student's-t TCC
Confidence Interval, Single Population Mean, Population Standard Deviation Unknown, Student-t is part of the collection col10555 written by Barbara Illowsky and Susan Dean with contributions from Roberta Bloom.

In practice, we rarely know the population **standard deviation**. In the past, when the sample size was large, this did not present a problem to statisticians. They used the sample standard deviation s as an estimate for σ and proceeded as before to calculate a **confidence interval** with close enough results. However, statisticians ran into problems when the sample size was small. A small sample size caused inaccuracies in the confidence interval.

William S. Gossett (1876-1937) of the Guinness brewery in Dublin, Ireland ran into this problem. His experiments with hops and barley produced very few samples. Just replacing σ with s did not produce accurate results when he tried to calculate a confidence interval. He realized that he could not use a normal distribution for the calculation; he found that the actual distribution depends on the sample size. This problem led him to "discover" what is called the **Student's-t distribution**. The name comes from the fact that Gosset wrote under the pen name "Student."

Up until the mid 1970s, some statisticians used the **normal distribution** approximation for large sample sizes and only used the Student's-t distribution for sample sizes of at most 30. With the common use of graphing calculators and computers, the practice is to use the Student's-t distribution whenever s is used as an estimate for σ.

If you draw a simple random sample of size n from a population that has approximately a normal distribution with mean μ and unknown population standard deviation σ and calculate the t-score $\mathrm{t} = \frac{\mathrm{x}-\mu}{\left(\frac{s}{\sqrt{n}}\right)}$, then the t-scores follow a **Student's-t distribution with $n-1$ degrees of freedom**. The t-score has the same interpretation as the **z-score**. It measures how far x is from its mean μ. For each sample size n, there is a different Student's-t distribution.

The **degrees of freedom**, $n - 1$, come from the calculation of the sample standard deviation s. In an earlier chapter, we used n deviations $(x - x \ \text{values})$ to calculate s. Because the sum of the deviations is 0, we can find the last deviation once we know the other $n - 1$ deviations. The other $n - 1$ deviations can change or vary freely. **We call the number $n - 1$ the degrees of freedom (df).**

Properties of the Student's-t Distribution

- The graph for the Student's-t distribution is similar to the Standard Normal curve.
- The mean for the Student's-t distribution is 0 and the distribution is symmetric about 0.
- The Student's-t distribution has more probability in its tails than the Standard Normal distribution because the spread of the t distribution is greater than the spread of the Standard Normal. So the graph of the Student's-t distribution will be thicker in the tails and shorter in the center than the graph of the Standard Normal distribution.
- The exact shape of the Student's-t distribution depends on the "degrees of freedom". As the degrees of freedom increases, the graph Student's-t distribution becomes more like the graph of the Standard Normal distribution.
- The underlying population of individual observations is assumed to be normally distributed with unknown population mean μ and unknown population standard deviation σ. The size of the underlying population is generally not relevant unless it is very small. If it is bell shaped (normal) then the assumption is met and doesn't need discussion. Random sampling is assumed but it is a completely separate assumption from normality.

A probability table for the Student's-t distribution can be used here. The table gives t-scores that correspond to the confidence level (column) and degrees of freedom (row). When using t-table, note that some tables are formatted to show the confidence level in the column headings, while the column headings in some tables may show only corresponding area in one or both tails.

A Student's-t table (See the Table of Contents **15. Tables**) gives t-scores

given the degrees of freedom and the right-tailed probability. The table is very limited. **Calculators and computers can easily calculate any Student's-t probabilities.**

The notation for the Student's-t distribution is (using T as the random variable) is

- $T \sim t_{\mathrm{df}}$ where $\mathrm{df} = n - 1$.
- For example, if we have a sample of size n=20 items, then we calculate the degrees of freedom as df=n−1=20−1=19 and we write the distribution as $T \sim t_{19}$

If the population standard deviation is not known, the **error bound for a population mean** is:

- $\mathrm{EBM} = t \cdot \left(\frac{s}{\sqrt{n}}\right)$
- t is the t-score.
- use $\mathrm{df} = n - 1$ degrees of freedom
- s = sample standard deviation

The format for the confidence interval is:

$(x - \mathrm{EBM}, x + \mathrm{EBM})$.

The TI-83, 83+ and 84 calculators have a function that calculates the confidence interval directly. To get to it,
Press `STAT`
Arrow over to `TESTS`.
Arrow down to `8:TInterval` and press `ENTER` (or just press `8`).

Example:
Exercise:

Problem:

Suppose you do a study of acupuncture to determine how effective it is in relieving pain. You measure sensory rates for 15 subjects with the results given below. Use the sample data to construct a 95% confidence interval for the mean sensory rate for the population (assumed normal) from which you took the data.

The solution is shown step-by-step and by using the TI-83, 83+ and 84+ calculators.
8.6 9.4 7.9 6.8 8.3 7.3 9.2 9.6 8.7 11.4 10.3 5.4 8.1 5.5 6.9

Solution:

- You can use technology to directly calculate the confidence interval.
- The first solution is step-by-step (Solution A).
- The second solution uses the Ti-83+ and Ti-84 calculators (Solution B).

Solution A
To find the confidence interval, you need the sample mean, x, and the EBM.

$x = 8.2267 \qquad s = 1.6722 \qquad n = 15$

$\mathrm{df} = 15 - 1 = 14$

$t = 2.14$ using the t-table.

$$\mathrm{EBM} = t \cdot \left(\frac{s}{\sqrt{n}}\right)$$

$$\mathrm{EBM} = 2.14 \cdot \left(\frac{1.6722}{\sqrt{15}}\right) = 0.924$$

$$x - \mathrm{EBM} = 8.2267 - 0.9240 = 7.3$$

$$x + \mathrm{EBM} = 8.2267 + 0.9240 = 9.15$$

The 95% confidence interval is **(7.30, 9.15)**.

We estimate with 95% confidence that the true population mean sensory rate is between 7.30 and 9.15.

Solution B
Using a function of the TI-83, TI-83+ or TI-84 calculators:

Press `STAT` and arrow over to `TESTS`.
Arrow down to `8:TInterval` and press `ENTER` (or you can just press `8`). Arrow to `Data` and press `ENTER`.
Arrow down to `List` and enter the list name where you put the data.
Arrow down to `Freq` and enter 1.
Arrow down to `C-level` and enter .95
Arrow down to `Calculate` and press `ENTER`.
The 95% confidence interval is (7.3006, 9.1527)
**With contributions from Roberta Bloom

Glossary

Confidence Interval (CI)
: An interval estimate for an unknown population parameter. This depends on:

- The desired confidence level.
- Information that is known about the distribution (for example, known standard deviation).
- The sample and its size.

Confidence Level (CL)
: The percent expression for the probability that the confidence interval contains the true population parameter. For example, if the $\mathrm{CL} = 90\%$, then in 90 out of 100 samples the interval estimate will enclose the true population parameter.

Degrees of Freedom (df)

The number of objects in a sample that are free to vary.

Error Bound for a Population Mean (EBM)

The margin of error. Depends on the confidence level, sample size, and known or estimated population standard deviation.

Normal Distribution

A continuous random variable (RV) with pdf $\mathrm{f(x)} = \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}$, where μ is the mean of the distribution and σ is the standard deviation. Notation: X ~ $N(\mu, \sigma)$. If $\mu = 0$ and $\sigma = 1$, the RV is called **the standard normal distribution**.

Standard Deviation

A number that is equal to the square root of the variance and measures how far data values are from their mean. Notation: s for sample standard deviation and σ for population standard deviation.

Student's-**t** Distribution

Investigated and reported by William S. Gossett in 1908 and published under the pseudonym Student. The major characteristics of the random variable (RV) are:

- It is continuous and assumes any real values.
- The pdf is symmetrical about its mean of zero. However, it is more spread out and flatter at the apex than the normal distribution.
- It approaches the standard normal distribution as n gets larger.
- There is a "family" of t distributions: every representative of the family is completely defined by the number of degrees of freedom which is one less than the number of data.

10.4 Confidence Intervals: Confidence Interval for a Population Proportion TCC

Confidence Interval for a Population Proportion is part of the collection col10555 written by Barbara Illowsky and Susan Dean with contributions from Roberta Bloom.

During an election year, we see articles in the newspaper that state **confidence intervals** in terms of proportions or percentages. For example, a poll for a particular candidate running for president might show that the candidate has 40% of the vote within 3 percentage points. Often, election polls are calculated with 95% confidence. So, the pollsters would be 95% confident that the true proportion of voters who favored the candidate would be between 0.37 and 0.43 : $(0.40 - 0.03, 0.40 + 0.03)$.

Investors in the stock market are interested in the true proportion of stocks that go up and down each week. Businesses that sell personal computers are interested in the proportion of households in the United States that own personal computers. Confidence intervals can be calculated for the true proportion of stocks that go up or down each week and for the true proportion of households in the United States that own personal computers.

The procedure to find the confidence interval, the sample size, the **error bound,** and the **confidence level** for a proportion is similar to that for the population mean. The formulas are different.

How do you know you are dealing with a proportion problem? First, the underlying **distribution is binomial**. (There is no mention of a mean or average.) If X is a binomial random variable, where n = the number of trials and p = the probability of a success. To form a proportion, take X, the random variable for the number of successes and divide it by n, the number of trials (or the sample size). The random variable "p-hat") is that proportion,

P -hat$= \frac{X}{n}$

When n is large and p is not close to 0 or 1, we can use the **normal distribution** to approximate the binomial. For our class we use the idea that

the sampling distribution is normal if 1) the sample is random 2) n*p is greater than or equal to 10 and 3) n(1-p) is greater than or equal to 10.

The confidence interval has the form $(\text{p-hat} - \text{EBP}, \text{p-hat} + \text{EBP})$.

$\text{p-hat} = \frac{x}{n}$

p-hat = the **estimated proportion** of successes (p-hat is a **point estimate** for p, the true proportion)

x = the **number** of successes.

n = the size of the sample

The error bound for a proportion is

$\text{EBP} = z \cdot \sqrt{\frac{\text{p-hat}\cdot(\text{1-p-hat})}{n}}$

This formula is similar to the error bound formula for a mean, except that the "appropriate standard deviation" is different. For a mean, when the population standard deviation is known, the appropriate standard deviation that we use is $\frac{\sigma}{\sqrt{n}}$. For a proportion, the appropriate standard deviation is $\sqrt{\frac{p\cdot(\text{1-p})}{n}}$.

However, in the error bound formula, we use $\sqrt{\frac{\text{p-hat}\cdot(\text{1-p-hat})}{n}}$ as the standard deviation, instead of $\sqrt{\frac{p\cdot(\text{1-p})}{n}}$

In the error bound formula, the **sample proportion** p-hat **and is an estimate of the unknown population proportions** p . The estimated proportion p-hat is used because p is not known. p-hat is calculated from the data. p-hat is the estimated proportion of successes. *1* minus p-hat is the estimated proportion of failures.

Again, the confidence interval can only be used if the number of successes the number of failures are both greater than or equal to 10.

Note: For the normal distribution of proportions, the z-score formula is as follows.

$$z = \frac{\text{p-hat} - p}{\sqrt{\frac{p \cdot \text{(1-p)}}{n}}}$$

Example:
Exercise:

Problem:

Suppose that a market research firm is hired to estimate the percent of adults living in a large city who have cell phones. 500 randomly selected adult residents in this city are surveyed to determine whether they have cell phones. Of the 500 people surveyed, 421 responded yes - they own cell phones. Using a 95% confidence level, compute a confidence interval estimate for the true proportion of adults residents of this city who have cell phones.

Solution

- You can use technology to directly calculate the confidence interval.
- The first solution is step-by-step (Solution A).
- The second solution uses a function of the TI-83, 83+ or 84 calculators (Solution B).

Solution:

$n = 500$ x = the number of successes $= 421$

$\text{p-hat} = \frac{x}{n} = \frac{421}{500} = 0.842$

$\text{p-hat} = 0.842$ is the sample proportion; this is the point estimate of the population proportion.

$1 - \text{p-hat} = 1 - 0.842 = 0.158$

From earlier in this chapter, a 95% confidence level leads us to a z-critical value of 1.96.

$$\text{EBP} = 1.96 \cdot \sqrt{\frac{(0.842)\cdot(0.158)}{500}} = 0.032$$

$\text{p-hat} - \text{EBP} = 0.842 - 0.032 = 0.81$

$\text{p-hat} + \text{EBP} = 0.842 + 0.032 = 0.874$

The confidence interval for the true binomial population proportion is $(\text{p-hat} - \text{EBP}, \text{p-hat} + \text{EBP}) =(0.810, 0.874)$.

Interpretation
We estimate with 95% confidence that between 81% and 87.4% of all adult residents of this city have cell phones.

Explanation of 95% Confidence Level
95% of the confidence intervals constructed in this way would contain the true value for the population proportion of all adult residents of this city who have cell phones.

Solution:

Using a function of the TI-83, 83+ or 84 calculators:

Press `STAT` and arrow over to `TESTS`.
Arrow down to `A:1-PropZint`. Press `ENTER`.
Arrow down to x and enter 421.
Arrow down to n and enter 500.
Arrow down to `C-Level` and enter .95.
Arrow down to `Calculate` and press `ENTER`.
The confidence interval is (0.81003, 0.87397).

Example:
Exercise:

Problem:

For a class project, a political science student at a large university wants to estimate the percent of students that are registered voters. He surveys 500 students and finds that 300 are registered voters. Compute a 90% confidence interval for the true percent of students that are registered voters and interpret the confidence interval.

Solution:

- You can use technology to directly calculate the confidence interval.
- The first solution is step-by-step (Solution A).
- The second solution uses a function of the TI-83, 83+ or 84 calculators (Solution B).

Solution A
$x = 300$ and $n = 500$.

$\text{p-hat} = \frac{x}{n} = \frac{300}{500} = 0.600$

$1 - \text{p-hat} = 1 - 0.600 = 0.400$

Because we have a 90% confidence level, z = 1.645.

$\text{EBP} = 1.645 \cdot \sqrt{\frac{(0.60)\cdot(0.40)}{500}} = 0.036$

$\text{p-hat} - \text{EBP} = 0.60 - 0.036 = 0.564$

$\text{p-hat} + \text{EBP} = 0.60 + 0.036 = 0.636$

The confidence interval for the true binomial population proportion is $(\text{p-hat} - \text{EBP}, \text{p-hat} + \text{EBP}) = (0.564, 0.636)$.
Interpretation:

- We estimate with 90% confidence that the true percent of all students that are registered voters is between 56.4% and 63.6%.
- Alternate Wording: We estimate with 90% confidence that between 56.4% and 63.6% of ALL students are registered voters.

Explanation of 90% Confidence Level
90% of all confidence intervals constructed in this way contain the true value for the population percent of students that are registered voters.

Solution B
Using a function of the TI-83, 83+ or 84 calculators:

Press `STAT` and arrow over to `TESTS`.
Arrow down to `A:1-PropZint`. Press `ENTER`.
Arrow down to x and enter 300.
Arrow down to n and enter 500.
Arrow down to `C-Level` and enter .90.
Arrow down to `Calculate` and press `ENTER`.
The confidence interval is (0.564, 0.636).

Calculating the Sample Size n

If researchers desire a specific margin of error, then they can use the error bound formula to calculate the required sample size.

Example:
Suppose a mobile phone company wants to determine the current percentage of customers aged 50+ that use text messaging on their cell phone. How many customers aged 50+ should the company survey in order to be 90% confident that the estimated (sample) proportion is within 3 percentage points of the true population proportion of customers aged 50+ that use text messaging on their cell phone.

Solution
From the problem, we know that **EBP=0.03** (3%=0.03) and

$z = 1.645$ because the confidence level is 90%
However, in order to find n , we need to know the estimated (sample) proportion p-hat. But, we do not know p-hat yet. Since we multiply p-hat and (1-p-hat) together, we make them both equal to 0.5 because p'q'= (.5)(.5)=.25 results in the largest possible product. (Try other products: (.6)(.4)=.24; (.3)(.7)=.21; (.2)(.8)=.16 and so on). The largest possible product gives us the largest n. This gives us a large enough sample so that we can be 90% confident that we are within 3 percentage points of the true population proportion. To calculate the sample size n, use the formula and make the substitutions.
This gives $n = \frac{1.645^2(.5)(.5)}{.03^2}$=751.7
Round the answer to the next higher value. The sample size should be 752 cell phone customers aged 50+ in order to be 90% confident that the estimated (sample) proportion is within 3 percentage points of the true population proportion of all customers aged 50+ that use text messaging on their cell phone.
**With contributions from Roberta Bloom.

Glossary

Binomial Distribution
: A discrete random variable (RV) which arises from Bernoulli trials. There are a fixed number, n, of independent trials. "Independent" means that the result of any trial (for example, trial 1) does not affect the results of the following trials, and all trials are conducted under the same conditions. Under these circumstances the binomial RV X is defined as the number of successes in n trials. The notation is: X~$B(n, p)$. The mean is $\mu = \text{np}$ and the standard deviation is $\sigma = \sqrt{\text{npq}}$. The probability of exactly x successes in n trials is $P\left(X = x\right) = \binom{n}{x}p^x q^{n-x}$.

Confidence Interval (CI)

An interval estimate for an unknown population parameter. This depends on:

- The desired confidence level.
- Information that is known about the distribution (for example, known standard deviation).
- The sample and its size.

Confidence Level (CL)

The percent expression for the probability that the confidence interval contains the true population parameter. For example, if the $\mathrm{CL} = 90\%$, then in 90 out of 100 samples the interval estimate will enclose the true population parameter.

Error Bound for a Population Proportion(EBP)

The margin of error. Depends on the confidence level, sample size, and the estimated (from the sample) proportion of successes.

Normal Distribution

A continuous random variable (RV) with pdf $\mathrm{f}(\mathrm{x}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$, where μ is the mean of the distribution and σ is the standard deviation. Notation: $X \sim N(\mu, \sigma)$. If $\mu = 0$ and $\sigma = 1$, the RV is called **the standard normal distribution**.

10.5 Confidence Intervals: Summary of Formulas TCC
changed p' notation

Formula General form of a confidence interval

$(\text{lower value}, \text{upper value}) = (\text{point estimate} - \text{error bound}, \text{point estimate} + \text{error bound})$

FormulaTo find the error bound when you know the confidence interval

$\text{error bound} = \text{upper value} - \text{point estimate}$ OR
$\text{error bound} = \frac{\text{upper value} - \text{lower value}}{2}$

FormulaSingle Population Mean, Known Standard Deviation, Normal Distribution

Use the Normal Distribution for Means $\text{EBM} = z \cdot \frac{\sigma}{\sqrt{n}}$

The confidence interval has the format $(x - \text{EBM}, x + \text{EBM})$.

FormulaSingle Population Mean, Unknown Standard Deviation, Student's-t Distribution

Use the Student's-t Distribution with degrees of freedom $\text{df} = n - 1$. $\text{EBM} = \text{t}* \cdot \frac{s}{\sqrt{n}}$

FormulaSingle Population Proportion, Normal Distribution

Use the Normal Distribution for a single population proportion $\text{p-hat} = \frac{x}{n}$

$\text{EBP} = z \cdot \sqrt{\frac{\text{p-hat} \cdot (\text{1-p-hat})}{n}}$

The confidence interval has the format $(\text{p-hat} - \text{EBP}, \text{p-hat} + \text{EBP})$.

FormulaPoint Estimates

x is a point estimate for μ

$p\prime$ is a point estimate for ρ

s is a point estimate for σ

10.5 Confidence Intervals: Practice 1 TCC

Student Learning Outcomes

- The student will calculate confidence intervals for means when the population standard deviation is known.

Given

The mean age for all Foothill College students for a recent Fall term was 33.2. The population standard deviation has been pretty consistent at 15. Suppose that twenty-five Winter students were randomly selected. The mean age for the sample was 30.4. We are interested in the true mean age for Winter Foothill College students. (http://research.fhda.edu/factbook/FH_Demo_Trends/FoothillDemographicTrends.htm

Let $X =$ the age of a Winter Foothill College student

Calculating the Confidence Interval

Exercise:

Problem: $x =$

Solution:

30.4

Exercise:

Problem: n=

Solution:

25

Exercise:

Problem: 15=(insert symbol here)

Solution:

σ

Exercise:

Problem: Define the Random Variable, X, in words.

X =

Solution:

the mean age of 25 randomly selected Winter Foothill students

Exercise:

Problem: What is x estimating?

Solution:

μ

Exercise:

Problem: Is σ_x known?

Solution:

yes

Exercise:

Problem:

As a result of your answer to (4), state the exact distribution to use when calculating the Confidence Interval.

Solution:

Normal

Explaining the Confidence Interval

Construct a 95% Confidence Interval for the true mean age of Winter Foothill College students.

Exercise:

Problem: How much area is in both tails (combined)? $\alpha =$ ________

Solution:

0.05

Exercise:

Problem: How much area is in each tail? $\frac{\alpha}{2} =$ ________

Solution:

0.025

Exercise:

Problem: Identify the following specifications:

- **a** lower limit =
- **b** upper limit =
- **c** error bound =

Solution:

- **a**24.52
- **b**36.28
- **c**5.88

Exercise:

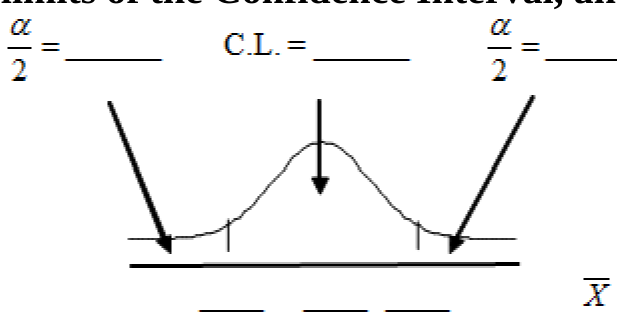
Problem: The 95% Confidence Interval is:___________________

Solution:

$(24.52, 36.28)$

Exercise:

Problem:
Fill in the blanks on the graph with the areas, upper and lower limits of the Confidence Interval, and the sample mean.



Exercise:

Problem: In one complete sentence, explain what the interval means.

Discussion Questions

Exercise:

Problem:

Using the same mean, standard deviation and level of confidence, suppose that n were 69 instead of 25. Would the error bound become larger or smaller? How do you know?

Exercise:

Problem:

Using the same mean, standard deviation and sample size, how would the error bound change if the confidence level were reduced to 90%? Why?

10.6 Confidence Intervals: Practice 2 TCC

Student Learning Outcomes

- The student will calculate confidence intervals for means when the population standard deviation is unknown.

Given

The following real data are the result of a random survey of 39 national flags (with replacement between picks) from various countries. We are interested in finding a confidence interval for the true mean number of colors on a national flag. Let $X =$ the number of colors on a national flag.

X	**Freq.**
1	1
2	7
3	18
4	7
5	6

Calculating the Confidence Interval

Exercise:

Problem: Calculate the following:

- **a** $x =$
- **b**$s_x =$
- **c**$n =$

Solution:

- **a**3.26
- **b**1.02
- **c**39

Exercise:

Problem:

Define the Random Variable, X, in words. $X =$
____

Solution:

the mean number of colors of 39 flags

Exercise:

Problem: What is x estimating?

Solution:

μ

Exercise:

Problem: Is σ_x known?

Solution:

No

Exercise:

Problem:

As a result of your answer to (4), state the exact distribution to use when calculating the Confidence Interval.

Solution:

t_{38}

Confidence Interval for the True Mean Number

Construct a 95% Confidence Interval for the true mean number of colors on national flags.

Exercise:

Problem: How much area is in both tails (combined)? $\alpha =$

Solution:

0.05

Exercise:

Problem: How much area is in each tail? $\frac{\alpha}{2} =$

Solution:

0.025

Exercise:

Problem: Calculate the following:

- **a**lower limit =

- **b**upper limit =
- **c**error bound =

Solution:

- **a** 2.93
- **b**3.59
- **c**0.33

Exercise:

Problem: The 95% Confidence Interval is:

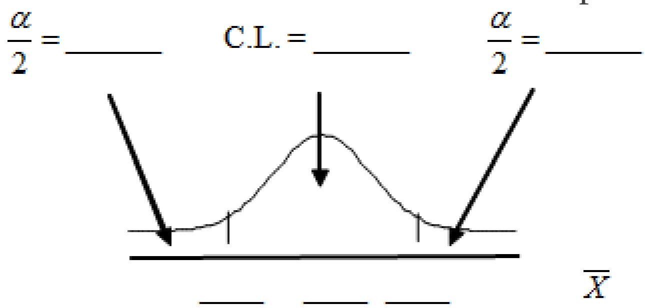
Solution:

2.93; 3.59

Exercise:

Problem:

Fill in the blanks on the graph with the areas, upper and lower limits of the Confidence Interval and the sample mean.



Exercise:

Problem: In one complete sentence, explain what the interval means.

Discussion Questions

Exercise:

Problem:

Using the same x, s_x, and level of confidence, suppose that n were 69 instead of 39. Would the error bound become larger or smaller? How do you know?

Exercise:

Problem:

Using the same x, s_x, and $n = 39$, how would the error bound change if the confidence level were reduced to 90%? Why?

10.7 Confidence Intervals: Practice 3 TCC

Student Learning Outcomes

- The student will calculate confidence intervals for proportions.

Given

The Ice Chalet offers dozens of different beginning ice-skating classes. All of the class names are put into a bucket. The 5 P.M., Monday night, ages 8 - 12, beginning ice-skating class was picked. In that class were 64 girls and 16 boys. Suppose that we are interested in the true proportion of girls, ages 8 - 12, in all beginning ice-skating classes at the Ice Chalet. Assume that the children in the selected class is a random sample of the population.

Estimated Distribution

Exercise:

Problem: What is being counted?

Exercise:

Problem: In words, define the Random Variable X. $X =$

Solution:

The number of girls, age 8-12, in the beginning ice skating class

Exercise:

Problem: Calculate the following:

- **a** $x =$
- **b** $n =$
- **c** p-hat $=$

Solution:

- **a**64
- **b**80
- **c**0.8

Exercise:

Problem: State the estimated distribution of X. X ~

Solution:

$B(80, 0.80)$

Exercise:

Problem: What is p-hat estimating?

Solution:

p

Exercise:

Problem: In words, define the Random Variable $P\prime$. $P\prime=$

Solution:

The proportion of girls, age 8-12, in the beginning ice skating class.

Exercise:

Problem: State the estimated distribution of $P\prime$. $P\prime$ ~

Explaining the Confidence Interval

Construct a 90% Confidence Interval for the true proportion of girls in the age 8 - 12 beginning ice-skating classes at the Ice Chalet.

Exercise:

Problem: How much area is in both tails (combined)? $\alpha =$

Solution:

.10

Exercise:

Problem: How much area is in each tail? $\frac{\alpha}{2} =$

Solution:

0.05

Exercise:

Problem: Calculate the following:

- **a**lower limit =
- **b**upper limit =
- **c**error bound =

Solution:

- **a**0.726
- **b**0.874
- **c**0.074

Exercise:

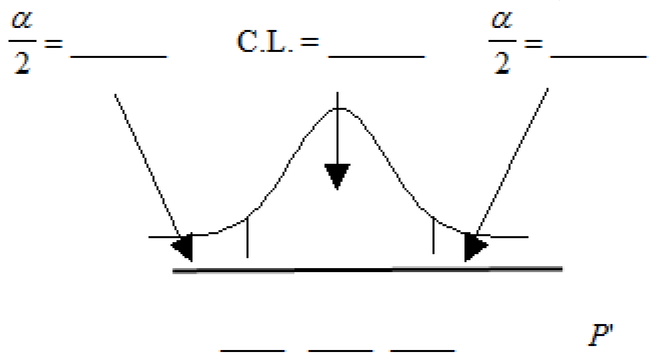
Problem: The 90% Confidence Interval is:

Solution:

(0.726; 0.874)

Exercise:

Problem:
Fill in the blanks on the graph with the areas, upper and lower limits of the Confidence Interval, and the sample proportion.



Exercise:

Problem: In one complete sentence, explain what the interval means.

Discussion Questions

Exercise:

Problem:

Using the same p-hat and level of confidence, suppose that n were increased to 100. Would the error bound become larger or smaller? How do you know?

Exercise:

Problem:

Using the same p-hat and $n = 80$, how would the error bound change if the confidence level were increased to 99%? Why?

Exercise:

Problem:

If you decreased the allowable error bound, why would the minimum sample size increase (keeping the same level of confidence)?

10.8 Confidence Intervals: Homework TCC

Note: If you are using a student's-t distribution for a homework problem below, you may assume that the underlying population is normally distributed. (In general, you must first prove that assumption, though.)

Exercise:

Problem:

Among various ethnic groups, the standard deviation of heights is known to be approximately 3 inches. We wish to construct a 95% confidence interval for the mean height of male Swedes. 48 male Swedes are surveyed. The sample mean is 71 inches. The sample standard deviation is 2.8 inches.

- **a**
  - **i**$x =$________
  - **ii**$\sigma =$ ________
  - **iii** $s_x =$________
  - **iv**$n =$________
  - **v**$n - 1 =$________
- **b**Define the Random Variables X and X, in words.
- **c**Which distribution should you use for this problem? Explain your choice.
- **d**Construct a 95% confidence interval for the population mean height of male Swedes.
  - **i**State the confidence interval.
  - **ii**Sketch the graph.
  - **iii**Calculate the error bound.

- **e**What will happen to the level of confidence obtained if 1000 male Swedes are surveyed instead of 48? Why?

Solution:

- **a**
  - **i**71
  - **ii**3
  - **iii**2.8
  - **iv**48
  - **v**47
- **c**$N\left(71, \frac{3}{\sqrt{48}}\right)$
- **d**
  - **i**CI: (70.15,71.85)
  - **iii**EB = 0.85

Exercise:

Problem:

In six packages of “The Flintstones® Real Fruit Snacks” there were 5 Bam-Bam snack pieces. The total number of snack pieces in the six bags was 68. We wish to calculate a 95% confidence interval for the population proportion of Bam-Bam snack pieces.

- **a**Define the Random Variables X and P', in words.
- **b**Which distribution should you use for this problem? Explain your choice
- **c**Calculate p'.
- **d**Construct a 95% confidence interval for the population proportion of Bam-Bam snack pieces per bag.
  - **i** State the confidence interval.
  - **ii**Sketch the graph.

- **iii**Calculate the error bound.

- **e**Do you think that six packages of fruit snacks yield enough data to give accurate results? Why or why not?

Exercise:

Problem:

A random survey of enrollment at 35 community colleges across the United States yielded the following figures (source: Microsoft Bookshelf): 6414; 1550; 2109; 9350; 21828; 4300; 5944; 5722; 2825; 2044; 5481; 5200; 5853; 2750; 10012; 6357; 27000; 9414; 7681; 3200; 17500; 9200; 7380; 18314; 6557; 13713; 17768; 7493; 2771; 2861; 1263; 7285; 28165; 5080; 11622. Assume the underlying population is normal.

- **a**
  - **i**$x =$
  - **ii** $s_x =$ ________
  - **iii**$n =$ ________
  - **iv**$n - 1 =$________
- **b**Define the Random Variables X and X, in words.
- **c**Which distribution should you use for this problem? Explain your choice.
- **d**Construct a 95% confidence interval for the population mean enrollment at community colleges in the United States.
  - **i**State the confidence interval.
  - **ii**Sketch the graph.
  - **iii**Calculate the error bound.
- **e**What will happen to the error bound and confidence interval if 500 community colleges were surveyed? Why?

Solution:

- **a**
  - **i**8629
  - **ii**6944
  - **iii**35
  - **iv**34
- **c** t_{34}
- **d**
  - **i**CI: (6244, 11,014)
  - **iii**EB = 2385
- **e**It will become smaller

Exercise:

Problem:

From a stack of IEEE Spectrum magazines, announcements for 84 upcoming engineering conferences were randomly picked. The mean length of the conferences was 3.94 days, with a standard deviation of 1.28 days. Assume the underlying population is normal.

- **a**Define the Random Variables X and X, in words.
- **b**Which distribution should you use for this problem? Explain your choice.
- **c**Construct a 95% confidence interval for the population mean length of engineering conferences.
  - **i**State the confidence interval.
  - **ii**Sketch the graph.
  - **iii**Calculate the error bound.

Exercise:

Problem:

Suppose that a committee is studying whether or not there is waste of time in our judicial system. It is interested in the mean amount of time individuals waste at the courthouse waiting to be called for service. The committee randomly surveyed 81 people. The sample mean was 8 hours with a sample standard deviation of 4 hours.

- **a**
  - **i**$x =$_________
  - **ii** $s_x =$ _________
  - **iii**$n =$_________
  - **iv**$n - 1 =$_________
- **b**Define the Random Variables X and X, in words.
- **c**Which distribution should you use for this problem? Explain your choice.
- **d**Construct a 95% confidence interval for the population mean time wasted.
  - **a**State the confidence interval.
  - **b**Sketch the graph.
  - **c**Calculate the error bound.
- **e**Explain in a complete sentence what the confidence interval means.

Solution:

- **a**
  - **i**8
  - **ii**4
  - **iii**81
  - **iv**80

- **c** t_{80}
- **d**

  - **i**CI: (7.12, 8.88)
  - **iii**EB = 0.88

Exercise:

Problem:

Suppose that an accounting firm does a study to determine the time needed to complete one person’s tax forms. It randomly surveys 100 people. The sample mean is 23.6 hours. There is a known standard deviation of 7.0 hours. The population distribution is assumed to be normal.

- **a**

  - **i**$x =$ ________
  - **ii**$\sigma =$________
  - **iii** $s_x =$________
  - **iv**$n =$ ________
  - **v**$n - 1 =$________

- **b**Define the Random Variables X and X, in words.
- **c**Which distribution should you use for this problem? Explain your choice.
- **d**Construct a 90% confidence interval for the population mean time to complete the tax forms.

  - **i**State the confidence interval.
  - **ii**Sketch the graph.
  - **iii**Calculate the error bound.

- **e**If the firm wished to increase its level of confidence and keep the error bound the same by taking another survey, what changes should it make?

- **f**If the firm did another survey, kept the error bound the same, and only surveyed 49 people, what would happen to the level of confidence? Why?
- **g**Suppose that the firm decided that it needed to be at least 95% confident of the population mean length of time to within 1 hour. How would the number of people the firm surveys change? Why?

Exercise:

Problem:

A sample of 16 small bags of the same brand of candies was selected. Assume that the population distribution of bag weights is normal. The weight of each bag was then recorded. The mean weight was 2 ounces with a standard deviation of 0.12 ounces. The population standard deviation is known to be 0.1 ounce.

- **a**
  - **i**$x =$ __________
  - **ii**$\sigma =$ __________
  - **iii** $s_x =$__________
  - **iv**$n =$ __________
  - **v**$n - 1 =$ __________
- **b**Define the Random Variable X, in words.
- **c**Define the Random Variable X, in words.
- **d**Which distribution should you use for this problem? Explain your choice.
- **e**Construct a 90% confidence interval for the population mean weight of the candies.
  - **i**State the confidence interval.
  - **ii**Sketch the graph.
  - **iii**Calculate the error bound.
- **f**Construct a 99% confidence interval for the population mean weight of the candies.

- **i**State the confidence interval.
- **ii**Sketch the graph.
- **iii**Calculate the error bound.

- **g**In complete sentences, explain why the confidence interval in (f) is larger than the confidence interval in (e).
- **h**In complete sentences, give an interpretation of what the interval in (f) means.

Solution:

- **a**
  - **i**2
  - **ii**0.1
  - **iii** 0.12
  - **iv**16
  - **v**15
- **b**the weight of 1 small bag of candies
- **c**the mean weight of 16 small bags of candies
- **d**$N\left(2, \frac{0.1}{\sqrt{16}}\right)$
- **e**
  - **i** CI: (1.96, 2.04)
  - **iii** EB = 0.04
- **f**
  - **i** CI: (1.94, 2.06)
  - **iii** EB = 0.06

Exercise:

Problem:

A pharmaceutical company makes tranquilizers. It is assumed that the distribution for the length of time they last is approximately normal. Researchers in a hospital used the drug on a random sample of 9 patients. The effective period of the tranquilizer for each patient (in hours) was as follows: 2.7; 2.8; 3.0; 2.3; 2.3; 2.2; 2.8; 2.1; and 2.4 .

- **a**
  - **i**$x =$ _________
  - **ii** $s_x =$ _________
  - **iii**$n =$ _________
  - **iv**$n - 1 =$ _________
- **b**Define the Random Variable X, in words.
- **c**Define the Random Variable X, in words.
- **d**Which distribution should you use for this problem? Explain your choice.
- **e**Construct a 95% confidence interval for the population mean length of time.
  - **i**State the confidence interval.
  - **ii**Sketch the graph.
  - **iii**Calculate the error bound.
- **f**What does it mean to be “95% confident” in this problem?

Exercise:

Problem:

Suppose that 14 children were surveyed to determine how long they had to use training wheels. It was revealed that they used them an average of 6 months with a sample standard deviation of 3 months. Assume that the underlying population distribution is normal.

- **a**

- **i** $x = $ _________
- **ii** $s_x = $ _________
- **iii** $n = $ _________
- **iv** $n - 1 = $ _________

- **b** Define the Random Variable X, in words.
- **c** Define the Random Variable X, in words.
- **d** Which distribution should you use for this problem? Explain your choice.
- **e** Construct a 99% confidence interval for the population mean length of time using training wheels.
  - **i** State the confidence interval.
  - **ii** Sketch the graph.
  - **iii** Calculate the error bound.
- **f** Why would the error bound change if the confidence level was lowered to 90%?

Solution:

- **a**
  - **i** 6
  - **ii** 3
  - **iii** 14
  - **iv** 13
- **b** the time for a child to remove his training wheels
- **c** the mean time for 14 children to remove their training wheels.
- **d** t_{13}
- **e**
  - **i** CI: (3.58, 8.42)
  - **iii** EB = 2.42

Exercise:

Problem:

Insurance companies are interested in knowing the population percent of drivers who always buckle up before riding in a car.

- **a**When designing a study to determine this population proportion, what is the minimum number you would need to survey to be 95% confident that the population proportion is estimated to within 0.03?
- **b**If it was later determined that it was important to be more than 95% confident and a new survey was commissioned, how would that affect the minimum number you would need to survey? Why?

Exercise:

Problem:

Suppose that the insurance companies did do a survey. They randomly surveyed 400 drivers and found that 320 claimed to always buckle up. We are interested in the population proportion of drivers who claim to always buckle up.

- **a**
  - **i**$x =$ ________
  - **ii**$n =$ ________
  - **iii**$p' =$ ________
- **b**Define the Random Variables X and P', in words.
- **c**Which distribution should you use for this problem? Explain your choice.
- **d**Construct a 95% confidence interval for the population proportion that claim to always buckle up.
  - **i**State the confidence interval.
  - **ii**Sketch the graph.
  - **iii**Calculate the error bound.

- **e**If this survey were done by telephone, list 3 difficulties the companies might have in obtaining random results.

Solution:

- **a**
  - **i**320
  - **ii** 400
  - **iii**0.80
- **c** $N\left(0.80, \sqrt{\frac{(0.80)(0.20)}{400}}\right)$
- **d**
  - **i** CI: (0.76, 0.84)
  - **iii** EB = 0.04

Exercise:

Problem:

Unoccupied seats on flights cause airlines to lose revenue. Suppose a large airline wants to estimate its mean number of unoccupied seats per flight over the past year. To accomplish this, the records of 225 flights are randomly selected and the number of unoccupied seats is noted for each of the sampled flights. The sample mean is 11.6 seats and the sample standard deviation is 4.1 seats.

- **a**
  - **i**$x =$ _________
  - **ii** $s_x =$ _________
  - **iii**$n =$ _________
  - **iv**$n - 1 =$ _________
- **b**Define the Random Variables X and X, in words.

- **c**Which distribution should you use for this problem? Explain your choice.
- **d**Construct a 90% confidence interval for the population mean number of unoccupied seats per flight.
    - **i**State the confidence interval.
    - **ii**Sketch the graph.
    - **iii** Calculate the error bound.

Exercise:

Problem:

According to a recent survey of 1200 people, 61% feel that the president is doing an acceptable job. We are interested in the population proportion of people who feel the president is doing an acceptable job.

- **a**Define the Random Variables X and P', in words.
- **b**Which distribution should you use for this problem? Explain your choice.
- **c**Construct a 90% confidence interval for the population proportion of people who feel the president is doing an acceptable job.
    - **i**State the confidence interval.
    - **ii**Sketch the graph.
    - **iii**Calculate the error bound.

Solution:

- **b** $N\left(0.61, \sqrt{\frac{(0.61)(0.39)}{1200}}\right)$
- **c**
    - **i**CI: (0.59, 0.63)
    - **iii** EB = 0.02

Exercise:

Problem:

A survey of the mean amount of cents off that coupons give was done by randomly surveying one coupon per page from the coupon sections of a recent San Jose Mercury News. The following data were collected: 20¢; 75¢; 50¢; 65¢; 30¢; 55¢; 40¢; 40¢; 30¢; 55¢; $1.50; 40¢; 65¢; 40¢. Assume the underlying distribution is approximately normal.

- **a**
  - **i**$x =$ _________
  - **ii** $s_x =$ _________
  - **iii**$n =$ _________
  - **iv**$n - 1 =$ _________
- **b**Define the Random Variables X and X, in words.
- **c**Which distribution should you use for this problem? Explain your choice.
- **d**Construct a 95% confidence interval for the population mean worth of coupons.
  - **i**State the confidence interval.
  - **ii**Sketch the graph.
  - **iii**Calculate the error bound.
- **e**If many random samples were taken of size 14, what percent of the confident intervals constructed should contain the population mean worth of coupons? Explain why.

Exercise:

Problem:

An article regarding interracial dating and marriage recently appeared in the Washington Post. Of the 1709 randomly selected adults, 315 identified themselves as Latinos, 323 identified themselves as blacks, 254 identified themselves as Asians, and 779 identified themselves as whites. In this survey, 86% of blacks said that their families would welcome a white person into their families. Among Asians, 77% would welcome a white person into their families, 71% would welcome a Latino, and 66% would welcome a black person.

- **a**We are interested in finding the 95% confidence interval for the percent of all black families that would welcome a white person into their families. Define the Random Variables X and P', in words.
- **b**Which distribution should you use for this problem? Explain your choice.
- **c**Construct a 95% confidence interval
    - **i**State the confidence interval.
    - **ii**Sketch the graph.
    - **iii**Calculate the error bound.

Solution:

- **b** $N\left(0.86, \sqrt{\frac{(0.86)(0.14)}{323}}\right)$
- **c**
    - **i**CI: (0.823, 0.898)
    - **iii** EB = 0.038

Exercise:

Problem: Refer to the problem above.

- **a**Construct three 95% confidence intervals.
  - **i**Percent of all Asians that would welcome a white person into their families.
  - **ii**Percent of all Asians that would welcome a Latino into their families.
  - **iii**Percent of all Asians that would welcome a black person into their families.
- **b**Even though the three point estimates are different, do any of the confidence intervals overlap? Which?
- **c**For any intervals that do overlap, in words, what does this imply about the significance of the differences in the true proportions?
- **d**For any intervals that do not overlap, in words, what does this imply about the significance of the differences in the true proportions?

Exercise:

Problem:

A camp director is interested in the mean number of letters each child sends during his/her camp session. The population standard deviation is known to be 2.5. A survey of 20 campers is taken. The mean from the sample is 7.9 with a sample standard deviation of 2.8.

- **a**
  - **i**$x =$ _________
  - **ii**$\sigma =$ _________
  - **iii** $s_x =$ _________
  - **iv**$n =$ _________
  - **v**$n - 1 =$ _________
- **b**Define the Random Variables X and X, in words.
- **c**Which distribution should you use for this problem? Explain your choice.

- **d**Construct a 90% confidence interval for the population mean number of letters campers send home.
  - **i**State the confidence interval.
  - **ii**Sketch the graph.
  - **iii**Calculate the error bound.
- **e**What will happen to the error bound and confidence interval if 500 campers are surveyed? Why?

Solution:

- **a**
  - **i** 7.9
  - **ii** 2.5
  - **iii** 2.8
  - **iv** 20
  - **v** 19
- **c** $N(7.9, \frac{2.5}{\sqrt{20}})$
- **d**
  - **i** CI: (6.98, 8.82)
  - **iii** EB: 0.92

Exercise:

Problem:

Stanford University conducted a study of whether running is healthy for men and women over age 50. During the first eight years of the study, 1.5% of the 451 members of the 50-Plus Fitness Association died. We are interested in the proportion of people over 50 who ran and died in the same eight–year period.

- **a**Define the Random Variables X and P’, in words.

- **b**Which distribution should you use for this problem? Explain your choice.
- **c**Construct a 97% confidence interval for the population proportion of people over 50 who ran and died in the same eight–year period.
    - **i**State the confidence interval.
    - **ii**Sketch the graph.
    - **iii**Calculate the error bound.
- **d** Explain what a “97% confidence interval” means for this study.

Exercise:

Problem:

In a recent sample of 84 used cars sales costs, the sample mean was $6425 with a standard deviation of $3156. Assume the underlying distribution is approximately normal.

- **a**Which distribution should you use for this problem? Explain your choice.
- **b**Define the Random Variable X, in words.
- **c**Construct a 95% confidence interval for the population mean cost of a used car.
    - **i**State the confidence interval.
    - **ii**Sketch the graph.
    - **iii**Calculate the error bound.
- **d**Explain what a “95% confidence interval” means for this study.

Solution:

- **a** t_{83}
- **b**mean cost of 84 used cars
- **c**

- **i**CI: (5740.10, 7109.90)
- **iii** EB = 684.90

Exercise:

Problem:

A telephone poll of 1000 adult Americans was reported in an issue of Time Magazine. One of the questions asked was “What is the main problem facing the country?” 20% answered “crime”. We are interested in the population proportion of adult Americans who feel that crime is the main problem.

- **a**Define the Random Variables X and P', in words.
- **b**Which distribution should you use for this problem? Explain your choice.
- **c**Construct a 95% confidence interval for the population proportion of adult Americans who feel that crime is the main problem.
  - **i**State the confidence interval.
  - **ii**Sketch the graph.
  - **iii**Calculate the error bound.
- **d**Suppose we want to lower the sampling error. What is one way to accomplish that?
- **e**The sampling error given by Yankelovich Partners, Inc. (which conducted the poll) is ± 3%. In 1-3 complete sentences, explain what the ± 3% represents.

Exercise:

Problem:

Refer to the above problem. Another question in the poll was “[How much are] you worried about the quality of education in our schools?” 63% responded “a lot”. We are interested in the population proportion of adult Americans who are worried a lot about the quality of education in our schools.

1. Define the Random Variables X and P', in words.
2. Which distribution should you use for this problem? Explain your choice.
3. Construct a 95% confidence interval for the population proportion of adult Americans worried a lot about the quality of education in our schools.

    - **i**State the confidence interval.
    - **ii**Sketch the graph.
    - **iii**Calculate the error bound.

4. The sampling error given by Yankelovich Partners, Inc. (which conducted the poll) is ± 3%. In 1-3 complete sentences, explain what the ± 3% represents.

Solution:

- **b** $N\left(0.63, \sqrt{\frac{(0.63)(0.37)}{1000}}\right)$
- **c**

    - **i**CI: (0.60, 0.66)
    - **iii** EB = 0.03

Exercise:

Problem:

Six different national brands of chocolate chip cookies were randomly selected at the supermarket. The grams of fat per serving are as follows: 8; 8; 10; 7; 9; 9. Assume the underlying distribution is approximately normal.

- **a**Calculate a 90% confidence interval for the population mean grams of fat per serving of chocolate chip cookies sold in supermarkets.
  - **i**State the confidence interval.
  - **ii**Sketch the graph.
  - **iii**Calculate the error bound.
- **b**If you wanted a smaller error bound while keeping the same level of confidence, what should have been changed in the study before it was done?
- **c**Go to the store and record the grams of fat per serving of six brands of chocolate chip cookies.
- **d**Calculate the mean.
- **e**Is the mean within the interval you calculated in part (a)? Did you expect it to be? Why or why not?

Exercise:

Problem:

A confidence interval for a proportion is given to be (– 0.22, 0.34). Why doesn’t the lower limit of the confidence interval make practical sense? How should it be changed? Why?

Try these multiple choice questions.

The next three problems refer to the following: According to a Field Poll, 79% of California adults (actual results are 400 out of 506 surveyed) feel that “education and our schools” is one of the top issues facing

California. We wish to construct a 90% confidence interval for the true proportion of California adults who feel that education and the schools is one of the top issues facing California. (Source: http://field.com/fieldpollonline/subscribers/)

Exercise:

Problem:A point estimate for the true population proportion is:

- **A**0.90
- **B**1.27
- **C**0.79
- **D**400

Solution:

C

Exercise:

Problem:A 90% confidence interval for the population proportion is:

- **A**(0.761, 0.820)
- **B**(0.125, 0.188)
- **C**(0.755, 0.826)
- **D**(0.130, 0.183)

Solution:

A

Exercise:

Problem: The error bound is approximately

- **A**1.581
- **B**0.791

- **C**0.059
- **D**0.030

Solution:

D

The next two problems refer to the following:

A quality control specialist for a restaurant chain takes a random sample of size 12 to check the amount of soda served in the 16 oz. serving size. The sample mean is 13.30 with a sample standard deviation of 1.55. Assume the underlying population is normally distributed.

Exercise:

Problem:

Find the 95% Confidence Interval for the true population mean for the amount of soda served.

- **A**(12.42, 14.18)
- **B**(12.32, 14.29)
- **C**(12.50, 14.10)
- **D**Impossible to determine

Solution:

B

Exercise:

Problem: What is the error bound?

- **A**0.87
- **B**1.98
- **C**0.99
- **D**1.74

Solution:

C

Exercise:

Problem:

What is meant by the term “90% confident” when constructing a confidence interval for a mean?

- **A**If we took repeated samples, approximately 90% of the samples would produce the same confidence interval.
- **B**If we took repeated samples, approximately 90% of the confidence intervals calculated from those samples would contain the sample mean.
- **C**If we took repeated samples, approximately 90% of the confidence intervals calculated from those samples would contain the true value of the population mean.
- **D**If we took repeated samples, the sample mean would equal the population mean in approximately 90% of the samples.

Solution:

C

The next two problems refer to the following:

Five hundred and eleven (511) homes in a certain southern California community are randomly surveyed to determine if they meet minimal earthquake preparedness recommendations. One hundred seventy-three (173) of the homes surveyed met the minimum recommendations for earthquake preparedness and 338 did not.

Exercise:

Problem:

Find the Confidence Interval at the 90% Confidence Level for the true population proportion of southern California community homes meeting at least the minimum recommendations for earthquake preparedness.

- **A**(0.2975, 0.3796)
- **B**(0.6270, 6959)
- **C**(0.3041, 0.3730)
- **D**(0.6204, 0.7025)

Solution:

C

Exercise:

Problem:

The point estimate for the population proportion of homes that do not meet the minimum recommendations for earthquake preparedness is:

- **A**0.6614
- **B**0.3386
- **C**173
- **D**338

Solution:

A